VISUALISING THE SOUNDFIELD AND SOUNDSCAPE: EXTENDING MACAULAY AND CRERAR'S 1998 METHOD

Iain McGregor, Alison Crerar, David Benyon and Grégory Leplâtre

School of Computing,
Napier University,
Edinburgh, UK.
{i.mcgregor, a.crerar, d.benyon, g.leplatre}@napier.ac.uk

ABSTRACT

The introduction of effective auditory warnings into a shared environment requires a prior understanding of the existing soundfield and soundscape. Reifying the physical and perceptual auditory environment enables a form of pre-auditioning, as well as the evaluation of any auditory augmentation. This paper describes the development of a visualisation technique for soundscape mapping. Building on earlier published work in sound classification, we report data captured using eighteen participants in a shared office environment. The resulting sound classification is used as the basis of a pictorial soundscape and soundfield visualisation. We show how this representation can be used to model the experiences of individuals, as well as subsets of users of the space.

1. INTRODUCTION

This paper addresses a method of visualising soundfields and soundscapes through the extension of Macaulay and Crerar's 1998 method [1]. Having applied Macaulay and Crerar's classification in isolation [2], and analyzed office inhabitant interviews [3], a number of missing elements were identified that would be necessary in order to successfully reify a shared auditory environment. An extended Macaulay and Crerar method (MC+) was developed, to capture these additional aspects of the soundscape, which involved four stages: capture, measurement, classification and visualisation. This paper describes how the MC+ method was applied to a shared office environment, the results of which are discussed below.

2. BACKGROUND

A single location, Napier University's School of Computing office, was chosen to test the MC+ method. The office was open-plan, with six *regular* (permanent) *inhabitants* and sixty-five potential *intermittent* inhabitants on the staff, and hundreds of potential *new* inhabitants in the form of students and visitors. A typical thirty minute time period was chosen for the study, specifically 2.45-3.15 on a Wednesday afternoon, and the MC+ method was applied as explained below. This mid-afternoon time slot was chosen as the office would have full *regular* occupancy, representative events would be taking place, and unlike the start or finish of the day, personal conversations would be less prevalent, and students would not be arriving en masse to hand in coursework.

A more complex environment was chosen than the academics' offices used when testing Macaulay and Crerar's original method [2]. This allowed for a broader range of sound events due to the complex nature of the activities conducted within the open-plan school office. There was a balance of *regular* inhabitants working at stationary positions, with continually changing *intermittent* inhabitants who provided a more dynamic atmosphere, with subjects moving about and entering and exiting more regularly. The office acts as the administrative centre for both staff and over 1000 students. Staff members mingle freely in the office, whereas students are confined behind a counter. There are six desks occupied by the regular office staff; they have conversations with each other and with academic staff who visit the room. This environment is typical of an open-plan office with regular visitors.

3. PARTICIPANTS

Eighteen participants were chosen to take part in the study. The number was based on there being six *regular* inhabitants, so an equal number of *intermittent* and new *inhabitants* were recruited. All the *regular* inhabitants had been working in this office environment for at least one year. They spent virtually all their time in this room so were very familiar with the environment. The six *intermittent* inhabitants were selected from academics within the School of Computing, who responded to e-mail requests. All intermittent users had more than one year's experience of visiting the administrative office, typically daily. Usually visits lasted anything from 5 minutes to a half-hour depending on the purpose.

The six *new* inhabitants were also Napier University employees, again selected from e-mail responses. In this case, no one from the group had ever set foot in the School of Computing office, nor had they any idea of its layout, beyond the map that was used for notation purposes. They had also never met any of the office's *regular* or *intermittent* inhabitants, so were unable to identify individuals within the recording. Each person had office space of their own, and was familiar with other school offices, and the type of interactions and sound sources contained within such environments.

4. METHOD

The MC+ method involved four stages: *capture*, *measurement*, *classification* and *visualisation*, as shown in Figure 1. The *capture* of the auditory environment through the creation of a *floor plan* and an *audio recording* enabled a consistent soundfield to be experienced by all of the *intermittent* and new *inhabitants*; and was also essential for accurate *notation* and *measurement* by the first author. The *classification* provided details about the perceived soundscape from the participants' perspectives. The *visualisation* was used as an accessible method of interpreting the results from individual participants as well as their respective groups. We explain each of these phases in the following subsections.

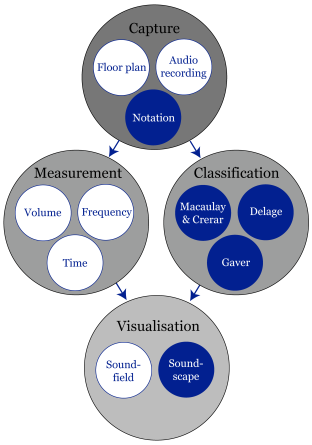

Figure1. *Extended Macaulay and Crerar Method (MC+).*

4.1. Capture

The capture stage involved the creation of a floor plan, followed by a surround sound multitrack recording, which was subsequently used in order to notate all of the audible sound events.

4.1.1. Floor plan

Firstly, a floor plan was created. This involved the measurement of the room and all fixed objects, such as desks, filing cabinets, windows and doors. These measurements were converted onto a floor plan with a scale of 100:1, which was overlaid with a grid of cells, each representing 50cm by 50cm. One row of additional cells was added around the perimeter, in order to allow the notation of sounds originating outside the room The grid was numbered in the same manner as an ordnance survey map, with 0 0 in the bottom left. Thus the room being modelled occupies 1 1 (bottom left) to 20 17 (top right), as shown in Figure 2.

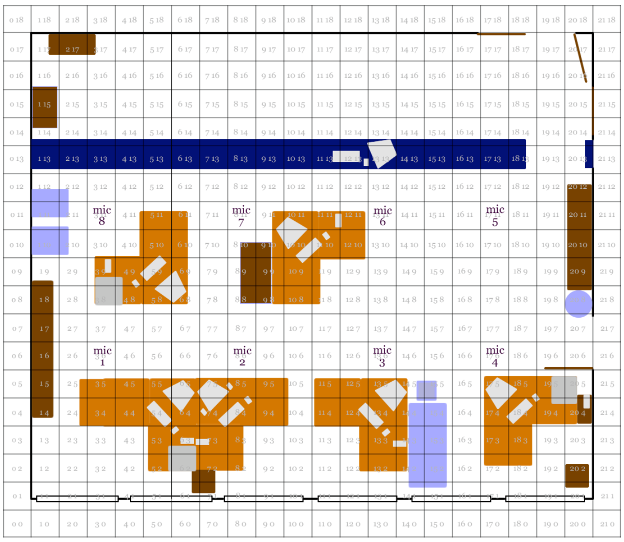

Figure 2. *Floor plan of School of Computing Office with fixed objects (bookcases, computers, desks, filing cabinets, water dispenser, worktop) and microphone positions (mic 1-8).*

4.1.2. Recording

For this study, an eight-channel, surround sound recording was made. Eight omni-directional microphones were mounted on stands at 1.5m in height. These were positioned as shown in Figure 2 (mics. 1-8); each microphone was mounted with a windshield. The choice of windshield, which completely enclosed the microphone suspending it from the support, enabled an element of shock mounting, which reduced the effect of any vibrations from the floor, or knocks to the stands. Notes about individual sound events were taken during the recording process to aid later identification.

The recording was made in a single thirty-minute pass onto eight separate channels. A separate eight channel pre-amplifier was used to minimize distortion and ensure consistency in both dynamics and frequency. Each channel was recorded uncompressed at 96kHz and 24 bits, which provided a theoretical dynamic range of c. 144 dB and ensured that the full audible range was covered. The high sampling rate meant that not only could ultrasonic frequencies be recorded (ensuring that associated phase cancellation could be reproduced), but also that the short time delays, with an accuracy of c. fifteen microseconds, that we rely on in order to accurately locate sounds, could be reproduced (something which is not possible at the standard CD sampling rate of 44.1 kHz) [4].

The microphones were erected four hours before the recording to allow inhabitants to become familiar with them, as well as identifying whether they caused a physical obstruction. No announcement was made when the recording started or finished, and the microphones were left erected for a further half hour so that participants were unaware of when the recording took place. The intention was to acclimatize the participants, check recording levels and minimize the impact of the recording on the soundscape. Eight-channel recording was chosen as it allowed more accurate positioning of the sound sources after the event. For reproduction, eight compact monitors were supplemented by four sub-bass units, whilst bass transmission can normally be considered omnidirectional, the low SPLs (Sound Pressure Levels) made accurate positioning of low frequency sounds, such as people walking on hollow resonant floors, difficult. The use of four sub bass units solved this problem, achieving a more accurate representation than that normally associated with a 5.1 or 7.1 system, where sub bass is normally located in front of the listener. This also compensated for the reduced frequency transmission range associated with compact monitors.

4.1.3. Notation

After the recording, all of the audible individual sound events were notated by listening back to the recording using the surround sound system: this process took about one hour per three minutes of recorded sound. Schafer [5] noted that the use of a microphone inhibits the ‘cocktail party’ effect, and thus it was found that listening back to the recording allowed a more accurate impression of how the soundscape is heard by new inhabitants, and helped to reduce reliance on the real-time auditory interpretation. Note was made of *Event, Source, Start Time, End Time* and *Location*, see Table 1. Events were classified by the action, such as typing; sources were identified either by the person or the object generating the sound. The start and end points were notated in hours, minutes and seconds and were rounded up to the nearest second. Location was notated by grid reference, and any movement of the sound source was also captured, such as an individual walking while talking.

Event	Source	Start Time	End Time	Location
Radio playing	Radio	00:00:00	00:30:00	13 3
Traffic	Vehicles	00:00:00	00:30:00	00 0 - 21 0
Typing	Catherine	00:00:00	00:00:05	19 3
Typing	Leanne	00:00:00	00:00:10	11 9
Mouse clicks	Fiona	00:00:04	00:00:06	12 3

Table 1: Example Sound Event Notation.

During the same thirty minute time period a week later, the *regular* inhabitants, whilst remaining in their normal work location, were asked to list all the sound events they consciously heard in terms of the *event*, *source* and *description*, as well as recording the *location* of the source with reference to the grid representing the room. On this occasion, the time of each sound event was not noted, as it differed from the original period. Inhabitants were asked only to pause momentarily to make notes rather than stop and actively listen. This was to try to ensure that all of the sounds that the participants usually created during their work would still be present, and thus could be heard by both themselves and their colleagues. This did mean that a number of sound events were missed, but it did allow for object-oriented descriptions of the events, (describing the object rather than the sound event or cause). This approach also helped to prevent all the participants stopping at once, as not all of them heard every event, and naturally individuals were loath to notate every single occurrence of repetitive events.

The *regular* inhabitants had a wealth of auditory experience to draw upon when discussing their shared auditory environment. This was only partially the case when it came to the *intermittent* inhabitants, and entirely absent with the *new* inhabitants. With both of the latter groups, it would have been unusual for them to spend a continuous 30 minutes in the office, and as such, their experiences of the auditory environment were likely to differ. In order to provide each of the remaining participants with a similar experience they were exposed to the surround sound recording of the office.

The participants were asked to make notes with regards to *event*, *source*, *description* and *location* of everything they heard while listening to the audio reproduction of the office. The inhabitants' notations were not used for the resultant maps, but were used to provide evidence of sound dimensions that might have been missing from the MC+ method, as well as confirming the validity of the dimensions which were present.

4.2. Measurement

Once the first author had made notation of all of the individual sound events, where possible, recordings and SPL readings of individual sound events were taken, these measurements were utilized to create the soundfield part of the map. The soundfield represents the quantifiable attributes of a sound event, which are independent of inhabitants' 'perceptions'. Each sound event is however unique, and these measurements are only representative of a single instance, because the complex interaction of materials and other sound events will affect each occurrence. In order to isolate the sound source from the background, recordings were made in mono, using the built in microphone on the SPL meter. Careful attention was paid to proximity, to reduce the effect of colouration from the microphone either being too close or too far away.

In order to be able to calculate an approximate sound pressure level, A scale peak readings were taken for each event. This ensured that a wide variety of SPL readings could be derived from the recordings due to the peak being known. The SPL meter was mounted on a tripod and the distance from the source measured. Ensuring that the peak measured was at least 6dB above the background set the distance. A difference of 6dB meant that the sound event was double the volume of the auditory background. Knowing the distance from the source allowed a calculation to be made as to the SPL level at one metre. The formula:

$$dB\ SPL = Max\ SPL + (20 \times Log\ (distance1/distance2)) \quad (1)$$

was used with the awareness that reverberation often amplifies a sound after a certain distance, which varies according to frequency and location.

Frequency range was calculated by passing the recording of the sound source through a spectrogram and notating the lowest and highest frequencies within the 6dB difference above the auditory background. These notes were collated into: *Event, Source, Time (Start & Stop), Location, SPL* and *Frequency Range* (see Table 2), which were then grouped together into a candidate sound event list of *Event, Source, Time Period and Location*, suitable for questioning inhabitants.

Event	Source	Start Time	End Time	Location	dB A	Hz
Radio playing	Radio	00:00:00	00:30:00	13 3	40	100Hz - 7kHz
Traffic	Vehicles	00:00:00	00:30:00	00 0 - 21 0	66	20Hz - 5kHz
Typing	Catherine	00:00:00	00:00:05	19 3	54	800Hz - 10kHz
Typing	Leanne	00:00:00	00:00:10	11 9	56	800Hz - 10kHz
Mouse clicks	Fiona	00:00:04	00:00:06	12 3	40	3.5 - 16.2kHz

Table 2. Example Sound Event Measurements

4.3. Classification

In the case of the *regular* participants, the classification interviews were conducted during a time period which represented the original recording as closely as possible, and at the same location as the original recording (their open-plan office). Inhabitants were interviewed using a comprehensive sound event list, created by the first author who notated every audible event from the multi track audio recording. Three forms of classification were applied at this point: Macaulay/Crerar's *sound types*, *information categories* and *acoustical information* [1], Delage's interactive functions [6] and Gaver's interacting materials [7] (see Table 3).

Frustrated by the lack of appropriate auditory models for the interaction designer, Macaulay and Crerar [1] studied the work of Brewster [8], Feld [9], Gaver [7] and Truax [10] as a basis for formulating a soundscape classification more appropriate to the field of Human Computer Interaction (HCI). They started with the belief that sound reveals information by situating individuals inside their 360 degree environment, rather than light, which presents information in the front 180 degrees [11]. The resulting model provides interactive systems designers with a framework for classifying sounds, which is a preliminary step in the move away from today's visually saturated interfaces. Macaulay and Crerar proposed a method of classifying constituents of soundscapes based upon (i) sound type, (ii) information category and (iii) acoustical information.

In 1998 Bernard Delage collaborated with Heleen Engelen to arrange a 'Sound Design Day', this was by invitation only, and involved architects, acousticians, computer scientists, composers, electro acousticians, scenographers, sound and visual designers all of whom had sound design experience. Whilst debating the role of sound and ergonomy, specifically within the realm of auditory feedback provided by manual tools, the group developed a list for the interactive function of sounds [6]. Gaver advocates an ecological approach to classifying sounds according to their 'audible source attributes'. Solids, gases or liquids generate these sound events and complex sounds can be described by either 'temporal patterning, compound or hybrid sources'. The results can be reproduced in map form in order to illustrate the qualitative nature of the sound events we hear. Gaver acknowledged that this classification was not complete, citing the voice, electricity and fire as possible simple, additional, sonic event candidates, as well as any definitive classification of a source being 'somewhat questionable' due the qualitative nature of listening. This alignment of the physical actions with everyday language does give a form of eliciting psychoacoustical responses, with a high degree of potential granularity when we include patterned, compound and hybrid events [7].

Sound Type	Example
Music	Any type of identifiable music, radio/stereo
Speech	Conversation
Other known	Identifiable recognised sounds
Other unknown	Unidentifiable unrecognised sounds
Information Category	**Example**
Visible entities and events	The phone ringing
Hidden entities and events	The photocopier round the corner being used
Imagined entities and events	Something big is happening as it has gone quiet
Patterns of events/entities	Someone is batch copying a large document
Passing of time	It's nearly deadline time (because the shift change is happening)
Emotions	A person is unhappy (tapping or slamming)
Euclidean Position	Person is moving around you
Acoustical information	**Example**
Foreground	Computer beep to attract your attention.
Contextual	Door opening (Help you orient to the nature of your environment.)
Background	Whine of disk drive providing reassurance or information about the state of the world.
Interactive functions	**Example**
Warning	Be careful
Assisting	Don't forget
Incitement	I am ready you can use me
Monitoring	In hospitals, in industry
Reassurance	You did OK
Forgiving	Try it again you'll succeed in the end
Guiding	Pedestrians at a crossroad
Protecting	Your car or house
Relaxing	So that you perform better
Neutral	No relevant information
Noise	Unwanted
Interacting materials	**Example**
Impact	Door is slammed or object is dropped
Scraping	Pen writing, Paper Rustling
Other vibration	(specify)
Explosion	Car backfiring
Continuous aerodynamic	Wind
Other aerodynamic sound	(specify)
Dripping	Water from tap
Splashing	Washing up
Other liquid sound	(specify)

Table 3. Extended Macaulay and Crerar Classification (MC+).

Delage's interactive functions were extended through the addition of '*neutral*' and '*noise*' categories (see Table 3). 'Neutral' denoted that the sound event had no information content, whereas 'noise' denoted that the sound was unwanted, and therefore a pollutant. Gaver's classification was applied as it stood in order to extend Macaulay and Crerar's 'sound type'. It was hoped to expand on Gaver's use of 'other' with reference to vibrations, as the work progressed.

4.4. Visualisation

Visualisation has always been more in the realm of the specialist acoustician, with the majority of designers being confined to waveforms, peak program meters (PPM) and occasionally spectrograms, all of which concentrate on dynamics, even in the case of spectrograms, where the frequency and energy of the signal are plotted against time. All of the dimensions were displayed on a single map in the manner of an interactive geographic information system (GIS). A map was created for each participant, carefully marking their point of listening (POL), so that all of the participants' maps could be overlaid. The results were then collapsed for the *regular* inhabitants, as well as the *intermittent*, the *new* and finally the three groups *combined*, resulting in four maps. When collapsing the results, a simple majority system was employed. When no single classification had a majority then multiple classifications were listed. The maps could then be viewed either as an A4/3 printout of the overview showing all of the sound events overlaid, or displayed on a 1024 x 768 monitor in its interactive form, with visualisations which are animated concurrently with the multi-track recording.

A variety of techniques were employed for visualisation: the source was listed according to two letters with a key containing a plain text description. *SPL* was represented by size,: the louder the sound the greater the diameter, the diameter was directly proportional to the SPL (see Figure 3). *Frequency* was represented through colour with frequency directly mapped to the visible spectrum. 20 Hz as a pure *Blue* and 20 kHz a pure red, with all of the intermediate frequencies spaced appropriately in a logarithmic scale across the spectrum. The temporal dimension was represented only on the animated version of the map; objects appeared and disappeared, as they were present during the recording. This allowed scrolling to establish which sound events were present concurrently.

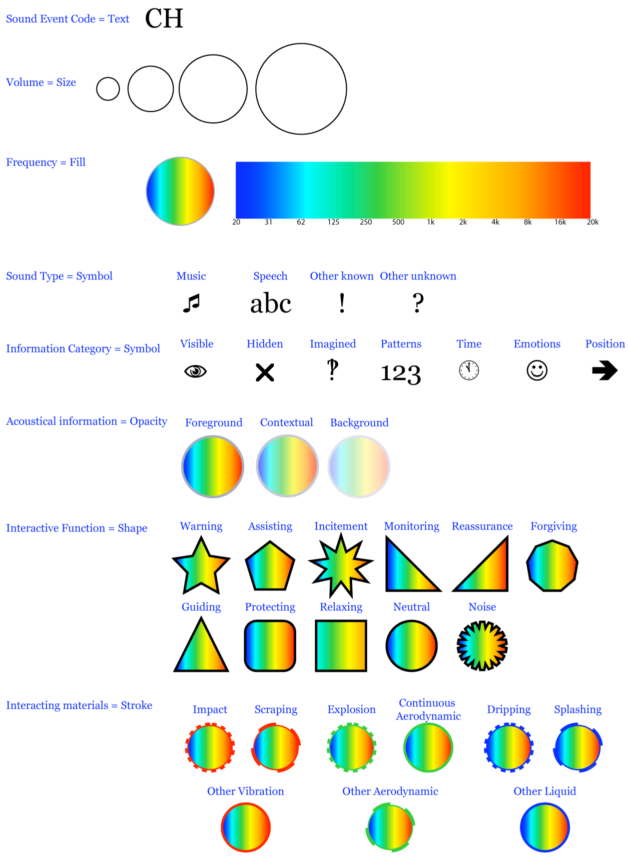

Figure 3. *Symbols used to create soundfield and soundscape maps.*

Sound type was represented through a symbol: *music* being a couple of quavers ♬; *speech* being the letters 'abc'; *other known* is an exclamation mark '!' and *other unknown* a question mark '?'. The information categories took the form of icons: *visible* was an eye 👁; *hidden* a barrier in the form of a cross ✕; *imagined* a question mark overlaid with an exclamation mark ‽. *Patterns* took the form of three consecutive numbers 123, with time being visualised by a clock face 🕓. A smiling face was used for *emotions* ☺, and a bold arrow for *position* ➔. Acoustical information was shown through opacity, with *foreground* having 100%, *contextual* 66% and *background* 33%.

Interactive function took the form of different shapes, the most common was *neutral*, which became a circle. *Noise*, *incitement* and *warning* were all stars, with 18, 9 and 5 points respectively. *Relaxing* and *protecting* were both squares, with the latter having rounded corners. *Monitoring* and *reassurance* became right-angled triangles in opposite directions, and the remaining classifications of *guiding*, *assisting* and *forgiving* became an isosceles triangle, pentagon and nonagon respectively.

The colour of the borders or strokes was used to denote the *interacting materials*. These could be grouped into three: *solid* (red), *liquid* (blue) and *gas* (green) each with a subcategory of *short* (short lines), *intermittent* (longer lines) or *continuous* (continuous line).

When a sound event fell between two or more classifications then multiple symbols were used for *sound type* and *information category*, with the shapes being split into half and rejoined to illustrate each *interactive function* or show different levels of *acoustical information*. An example of this can be seen in figures 5 and 6 where a drawer being closed (AX) is classified by new inhabitants as being *visible*, whereas the regular inhabitants considered it to be both *visible* and *hidden*. Finally, composite classifications of interacting materials were built up, for example, a solid green line denoting *continuous aerodynamic* could be surrounded by a series of short red lines signifying *impact*.

5. RESULTS

The results were examined in a number of ways: the notation produced by the first author from the surround sound recording was compared with the notations produced by the eighteen volunteers, and these were in turn compared against the MC+ classification list in order to assess how far it accommodated the observations. The next step was the measurement of each sound source by the first author followed by the classification of each sound event by each of the 18 participants, finishing with the visualisation of the results.

5.1. Capture

The capture stage involved the creation of a floor plan, an audio recording and notation of the auditory environment (as shown in Figure 1). This met with differing degrees of success as explained below.

5.1.1. Floor plan

The floor plan was extremely useful during the classification process as it allowed participants to recall different sound sources according to their personal orientation in the sound field. The grid size was appropriate at this stage, but when it came to notating the positions of sound sources, degrees within the cells would have given better accuracy.

5.1.2. Audio Recording

Only a single inhabitant referred to the recording equipment whilst the recording was taking place, and there was a single instance of mobile phone interference. Other than that the quality was sufficiently high to notate accurately all of the sound events. The multiple channels made it especially easy to cross reference notes made during the recording session with sound events audible on the recording, especially when it came to spatial notation. Participants uniformly commented about how immersive the experience was and how they often had to remind themselves that it was a recording.

5.1.3. Notation

The 30-minute recording contained 435 distinct sound events that were identified and notated by the researchers: these were emitted by 139 unique sound sources. These varied from stationary sources which were continuously audible, such as a radio playing, through to intermittent sources, an example of which was an individual talking while walking, as well as single events in the case of the water dispenser.

Concerning the inhabitants' notations, there was a large difference between the median number of events listed by the *regular* inhabitants when compared to the *intermittent* and *new* (see Table 4). As might be expected, due to familiarity and the effects of habituation, the *regular* inhabitants notated the fewest number of sources, whereas the *new* inhabitants, never having experienced the environment before recorded the highest by a factor of 7:1. The combined total of 136 distinct events reported by all 18 participants compared to the 435 originally notated, was partially due to only notating a sound source once and subsequently ignoring it, as well simple omission due to the participant not consciously hearing the event.

Group	Events	Descriptions	Grid ref	Orientation
Regular	11	100%	100%	0%
Intermittent	48	12%	72%	4%
New	72	17%	92%	0%
Combined	43	27%	98%	0%

Table 4. Median of inhabitants' notations, specifying number of events recorded, percentage of descriptions, grid references and orientation for sound events.

The number of descriptions provided of individual sound events, beyond source and action, did not change significantly by group, but the change in percentage was evident. Two factors were at work here, almost all of the participants found it hard to describe the sound event beyond stating the source and the event, but also the concern was to notate as many as possible, and the descriptions were deemed not as important as the source identification. The incidence of participants providing a grid reference was considerably higher, showing that the ability to localize the source, even without a visual cue was high, although not always possible. A few of the intermittent inhabitants preferred to indicate a sound source's orientation in relation to their own location.

In terms of the individual types of sound notated: speech was by far the most common, followed by typing, the radio, walking, printing and a water dispenser. These six were all referred to by at least 78% of the respondents and represented 54% of the total instances. All of which with the possible exception of the radio are generic to offices, and easily identifiable. A further ten sound events, which represented 30% of the total instances, were recorded by between 50-72% of the inhabitants. Only 2% of the total instances, such as handling a book, and dropping a bag, were noticed only by individuals. Therefore, 84% of the total sound instances detailed by the participants were captured by at least 50% of the respondents, illustrating a level of consistency across the three groups.

5.2. Measurement

All of the sound sources were measured from the original sources, rather than relying on the multitrack recording. The highest recorded SPL was a window being closed at 68 dB A, and the quietest was a person stapling at only 31 dB A. Speech fell between 36-64 dB A, depending upon context. This was very surprising as they were all comparatively low levels, with only 15% of the sound sources being above 60 dB A, and 45% being below 50 dB A.

When measuring the frequency range, almost the full audible frequency range was present. 20 Hz was generated by the traffic, and 19.6 kHz was present in the form of harmonics when some coins were dropped on a desk. The spectrograms were generated from recordings made direct to DAT through the SPL meter's built in microphone, and then replayed directly through the processor using peak capture, to generate the final figures.

Temporal measurement was made direct from the surround sound recording and rounded up to the nearest second. A single sound event would be identified from the recording and then listened to until it was no longer audible, the internal counter would then be used to calculate its length. More accurate timing was possible, even down to individual samples, but not applied at this point.

5.3. Classification

After the notation was created, participants were first asked if they were aware or unaware of a specific sound event. There was no significant difference between the three groups regarding the mean number of sound events heard (overall 59% of the sounds present were detected). This showed that the use of the surround sound recording provided a similar auditory experience to that of being in the real environment as the *regular* inhabitants were aware of almost an identical number of sound events as the *intermittent* and *new*.

5.3.1. Macaulay and Crerar's Sound Type, Information Category and Acoustical Information

Within the sound type classification it became apparent that the majority of the heard sound events were *other known* (71%). Ballas raises concerns about classifications, which are generated "by exclusion", in this case 'other known' and 'other unknown' [12]. From preliminary findings it would appear that replacing 'other known' with 'environmental' would be more appropriate, with 'other unknown', becoming simply 'unknown'. *Speech* was the next most common (25%) with only two percent of the sound sources being described as *unknown*, and none in the case of the *regular* inhabitants.

With regards to the *information category* the groups were again surprisingly similar. When questioning the *intermittent* and *new* inhabitants they were asked if they thought the specified sound event would be *visible* etc. if they were in the office. *Visible* was by far the most common with an average of 63% across all three groups, with little variation. The *new* group had a greater percentage of hidden (36%) compared to 11% with the *regular* group, who cited *emotions* and *position* (both 11%), which is understandable as they regularly use sound to gauge the emotions of their fellow inhabitants, as well as their positions within the environment, without having to look up from their desks.

With reference to the *acoustical information* there was more of a difference between the three groups. Regular inhabitants classified 49% of the sounds they heard as *background* and 38% as *foreground*, showing that a large proportion of sound events were irrelevant to their working lives. The *intermittent* group assigned 51% of the sounds they heard to the *contextual* category, illustrating their potential for interaction, whereas, *new* inhabitants produced a more even spread of observations across the three categories, with *background* being the most frequent (44%).

5.3.2. Delage's Interactive Functions and Gaver's Interacting Materials

Questioning the inhabitants of the soundscape about an individual sound event's interactive function gave us an insight into its perceived semantics. Not only was it possible to see where listeners shared interpretations, but also where there was a mismatch between the intended design of a sound event and its common interpretation, as in the case of the alert sound for the hand scanner, which was uniformly classed as noise. *Neutral* was the majority classification across all groups (54%) followed by *noise* (19%). Seven of the other classifications were fairly evenly spread except for *forgiving* and *protecting* which were unused. What was evident was the lack of any sound design within the office environment, but despite that, little of the sound was classed as *noise* despite the preponderance of sound from machinery such as printers and traffic, or doors opening and closing.

Gaver himself suggested extending his *interacting materials* to include *voice* and *fire* [7]. Inhabitants did not embrace the hybrid or compound aspects of this classification, preferring to use a single description; this again bears out Ballas' findings as to the preference for inhabitants describing semantic rather than acoustic properties, when questioned [12]. The two most common forms of classifications were *impact* (29%) and *continuous aerodynamic* (21%). There was little evidence of any *liquid* sounds, which corresponded with the lack of any liquids within the office, beyond the water dispenser. *Impact* was commonly used when classifying sounds associated with computer interaction such as typing and mouse use. *Continuous aerodynamic* was applied to the traffic and speech. The current nine types of *interacting materials* could be easily replaced by: solid, air and liquid each with single, intermittent and continuous states.

5.4. Visualisation

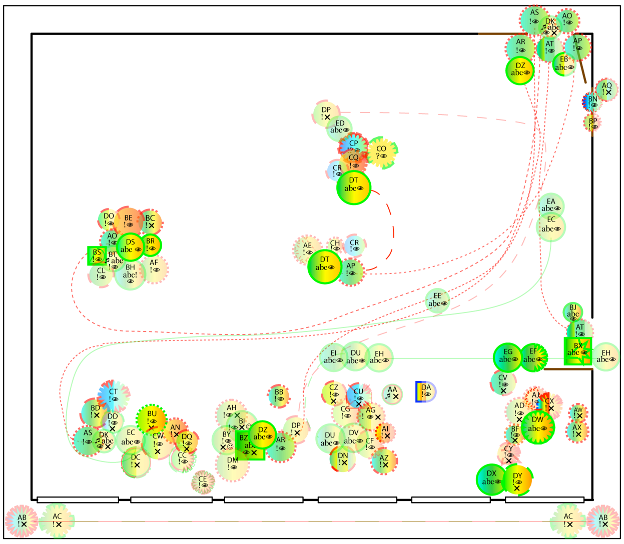

Figure 4. *Soundfield and Soundscape map representing new inhabitants' responses.*

A map was created for each participant, as well as each of the four groups: *regular*, *intermittent*, *new* and *combined* (for example, Figure 4). Only the walls, windows and doors were retained from the original floor plan, in order to allow easy identification of the spatial dimensions, otherwise only the sound events were included. The soundfield measurements remained a constant, the only factor, which affected their visualisation, was when individuals were unaware of the sound event, at which point it was omitted from the map.

On the macro scale it was easy to see the soundfield measurements and the classifications, especially the *acoustical information* (opacity) and *interactive function* (shape) (see Figure 4). The symbols and strokes required to be viewed at a larger scale as well as with closer attention, especially when they represented multiple classifications (Figure 5). Some of the symbols were more successful than others: *music*, *speech*, *other known*, *visible*, *time* and *emotions*, all have readily associated symbols, while the others have not: the choices were close enough to make them easy for the authors to remember, but probably not for others to guess.

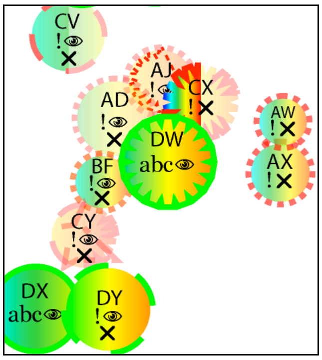

Figure 5. *Magnified area of Soundfield and Soundscape map representing new inhabitants' responses.*

The map was useful in establishing an overview, the majority of the sound events were *neutral*, and of a similar volume. *Noise* was easy to spot as were the relative interests of the group. For example, an area of the office was virtually ignored by one group but actively listened to by another (see Figures 5 and 6).

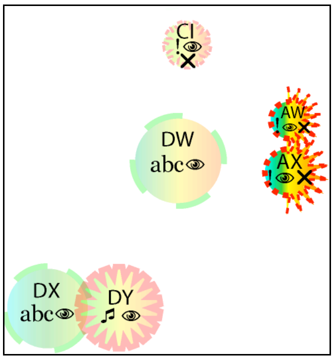

Figure 6. *Magnified area of Soundfield and Soundscape map representing regular inhabitants' responses (Identical coordinates to Figure 5).*

It was also clear that visitors to the office started their conversations upon entering the room, even if the person they wished to converse with was located at the far side. Surprisingly the *regular* inhabitants classed this as *neutral*, showing they were habituated to it, this was illustrated through the low level of opacity, which was easily visible on the macro scale, when viewing the entire map.

6. APPLICATIONS

There are a number of potential applications for this mapping method, both for the design and evaluation of auditory environments. The resultant maps can be utilized for design in a couple of ways. The first can be to recreate a soundscape in a virtual environment. When creating a virtual environment the designer might wish to control when events occur, so the map is an illustration of what took place during the time period under study. This allows the designer control over what they wish to include or exclude, in order to create a representative environment.

A second application is the design of auditory elements for augmentation of a real world auditory environment. The soundscape map shows what existing sounds a proposed sound would have to compete with, and allows new sounds to be overlaid onto the visualisation in order to see them in situ. The method also gives a mechanism for designing the sound itself. Designers can hear what inhabitants already consider to be aesthetically pleasing alert sounds, or any other combination within the classification, and design the new sounds accordingly. The proposed enhancements can then be tested by overlaying the sounds over the recording of the virtual soundscape and questioning the inhabitants.

The method can be used to evaluate soundscapes, sounds as well as to evaluate reproduction equipment. Both real and virtual soundscapes can be evaluated by following the method described and assessing the results. Sounds which are unwanted will be classified as *noise*, and sounds which do not serve the desired purpose due to being aesthetically displeasing, or masked will also be seen. With virtual soundscapes, the mapping process will illustrate how closely they match what the designer intended. This will be even more obvious if the designer used this method during the design stages. The comparison of different inhabitants' maps would illustrate the wide range of experiences, allowing successful sound events to remain, and unsuccessful ones to be re-designed.

Sounds could be evaluated by placing them within the virtual soundscape and seeing how they affect the results by applying the process described here. It will become plain if sounds are masked, or if the inhabitant[s] did not experience the sound in the way that the designer intended. This is easier if the designer first classifies the sounds themselves, thereby allowing direct comparison, as well as showing consistency of experience/interpretation with multiple inhabitants. This also allows contextual analysis, ensuring that the sound does not mask or contradict other sound events within the environment. The sound can also be placed into the physical environment and the method applied after a specified period to establish how it has altered the environment, and how it affects other inhabitants.

Reproduction equipment can be studied through comparison techniques. The perceived sound stage can be mapped in an identical manner using reference quality equipment or the original source, and then compared to the equipment under test. If the reproduction is 'coloured' in any way, such as through poor spatial definition the positioning of sound sources on the map will be vague, or the sound source will occupy a far greater area than intended. If there is frequency colouration then the notated colour of the sound events will change, if there is too much compression, all of the sound events will be a similar size.

When testing Head related Transfer Functions, if the reproduction is successful then the sounds should occupy a larger area than if they were In Head Located (IHL). Successful head tracking would have different participants placing the sound events in similar locations, with unsuccessful tracking resulting in divergent positions. Comparing the perceived brightness of individual sound events, as well as the number of separate events perceived could test data compression.

7. CONCLUSIONS

The university school office used for this experimental environment had six permanent inhabitants, and a large number of intermittent visitors (students and academic staff). The predominant sound was that of speech, with a single music source and a variety of known sounds such as traffic and typing. It was found that intermittent inhabitants have a considerable impact on the soundscape, starting conversations upon entering the room and continuing them while not even facing the person they are conversing with, and doing this irrespective of whether other inhabitants are disturbed by them. The practice of the 'out loud' is of benefit to the regular inhabitants, who make queries from one end of the room to the other, as well as voicing non-specific comments, which are intermittently listened to, sometimes causing laughter [1]. One *regular* inhabitant, through the use of a barely audible web radio station, customized the auditory environment around her desk, carving out a personal auditory space. This was the only instance of auditory personalization and was almost always masked by other sounds throughout the time period studied.

There was a dramatic disparity between the number of sounds that occurred and those reported by the participants, in some cases a ratio of 10:1. This can be partially explained by the fact that the participants are adept at relegating sounds to the 'background' through habituation. *Intermittent* inhabitants had the greatest effect on the soundscape through their desire to hold conversations across the room. However, the *regular* inhabitants are adept at altering the volume of their speech according to who is present. During the period under study, one participant became aware of how loud the street sounds were from an open window and closed it. During a preliminary observational session, the participants became aware of how loud a hinged countertop was (over 100 dB (A)), and subsequently stopped using it. Nevertheless, apart from these instances the regular inhabitants had little control over their auditory environment, and have shown little interest in wanting more control of their auditory environment.

Holman talks about a few additional factors, which are of interest to sound designers for film and television, notably radiation intensity and room absorption [13]. Radiation intensity refers to the fact that different frequencies are emitted with different intensities around an object; with often the fundamental frequency only emitted in a single direction, such as is the case with speech, where the ideal listening position is facing the speaker. This would be relatively easy to achieve by taking a variety of measurements of the sound source in situ. This would prove invaluable when deciding upon the placement of objects, both real and virtual, as an object with a greater range of radiation intensity is often easier to locate within an environment from the point of view of the listener. Room absorption can vary dramatically within a room, as well as from room to room. Normally rated on a scale of 0 to 1, with 1 representing 100% absorption, such as in an open window where none of the sound is reflected, and 0 representing 100% reflection, such as a perfect echo [14]. Scaling the auditory environment would allow us a level of predictability as to how new sounds would alter the soundfield. If we measured the room absorption by frequency, then the scale would become more accurate, informing us whether the room acoustics would react like an indoor swimming pool or other typical acoustic environments. It should be noted that this measure is different from reverberation time (RT), which measures the length of time over which a sound decays by a specific amount. RT measurements cannot be obtained at initial design stage, because they ideally require occupants to be present [15]. This is a problem faced by many designers, such as concert hall acousticians.

We have described a promising new approach to soundscape mapping (the MC+ method) and presented some preliminary results obtained by the researchers who created it. Clearly, the effectiveness of what we have produced needs testing by a variety of sound and interaction designers, who are the intended end users of the proposed mapping method.

8. REFERENCES

[1] C. Macaulay and A. Crerar, "'Observing' the Workplace Soundscape: Ethnography and Auditory Interface Design," presented at ICAD 98, International Conference on Auditory Display, University of Glasgow, Glasgow, 1998.

[2] I. McGregor, A. Crerar, D. Benyon, and G. Leplatre, "Workplace Soundscape Mapping: A Trial of Macaulay and Crerar's Method," presented at ICAD 2006, London, 2006.

[3] I. McGregor, A. Crerar, D. Benyon, and G. Leplatre, "Mapping Workplace Soundscapes: Reifying Office Auditory Environments," presented at World Forum for Acoustic Ecology, Hirosaki, Aomori, Japan, 2006.

[4] Y. A. Al'tman, N. N. Bekhterev, S. F. Vaitulevich, and N. I. Nikitin, "The Role of Phase Changes in Sound Signals in Localization of Sound Sources," Neuroscience and Behavioral Physiology, vol. 34, pp. 765-771, 2004.

[5] R. M. Schafer, "A l'ecoute de l'environnement," in Yearbook of Soundscape Studies 'Northern Soundscapes', Vol. 1, 1998, R. M. Schafer and H. Jarviluoma, Eds. Tampere: University of Tampere, Department of Folk Tradition, 1998, pp. 158-163.

[6] B. Delage, "On sound design," in Stockholm, Hey Listen!, From Awareness to Action, H. Karlsson, Ed. Stockholm: The Royal Swedish Academy of Music, 1998, pp. 67-73.

[7] W. W. Gaver, "What in the World do we Hear?," Ecological Psychology, vol. 5, pp. 1-29, 1993.

[8] S. A. Brewster, "Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces," in Human-Computer Interaction Group, Department of Computer Science. York: University of York, 1994, pp. 277.

[9] S. Feld, Sound and Sentiment: Birds, Weeping, Poetics and Song in Kaluli Expression. Philadelphia: University of Pennsylvania Press, 1990.

[10] B. Truax, Acoustic Communication, 2nd ed. Norwood: Ablex Publishing Corporation, 2001.

[11] W. Ong, The Presence of the Word: Some Prolegomena for Cultural and Religious History. New York: Simon & Schuster, 1967.

[12] J. A. Ballas and J. H. Howard Jr, "Interpreting the Language of Environmental Sounds," in Environment and Behaviour, vol. 19, 1987, pp. 91-114.

[13] T. Holman, Sound for Film and Television. Boston: Focal Press, 1997.

[14] G. M. Ballou, "Fundamentals and Units of Measurement," in Handbook for Sound Engineers, G. M. Ballou, Ed., 3rd ed. Boston: Focal Press, 2002, pp. 1491-1530.

[15] W. Ahnert and H.-P. Tennhardt, "Acoustics for Auditoriums and Concert Halls," in Handbook for Sound Engineers, G. M. Ballou, Ed. Boston: Focal Press, 2002, pp. 109-155.